

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-110751

(43) 公開日 平成7年(1995)4月25日

(51) Int. Cl. ⁴	識別記号	庁内整理番号	P I	技術表示箇所
G 0 6 F 3/14	3 4 0 B			
3/16	3 2 0 A	7165-5B		

審査請求 未請求 請求項の数 7 O L (全 10 頁)

(21) 出願番号 特願平5-254218

(22) 出願日 平成5年(1993)10月12日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 松浦 博

神奈川県川崎市幸区堀川町70番地 株式会社

東芝御町工場内

(72) 発明者 神尾 広幸

神奈川県川崎市幸区堀川町70番地 株式会社

東芝御町工場内

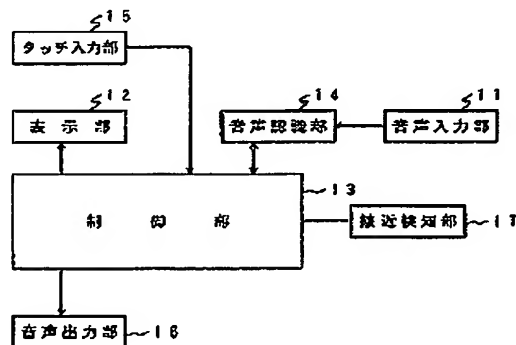
(74) 代理人 弁理士 鈴江 武彦

(54) 【発明の名称】 マルチモーダル装置

(57) 【要約】

【目的】 音声入力手段とタッチ入力手段との連携をとり、利用者の入力操作が容易且つ確実に行え、使い勝手が向上できるようにする。

【構成】 利用者の接近が接近検知部17により検知されると、制御部13は、表示部12に多数の入力項目を分類して表示し、音声入力部11からの音声による項目入力はいずれも受け付ける一方、タッチ入力部15からのタッチ入力は分類の代表項目だけを受け付けるようにする。代表項目の1つがタッチ入力されると、制御部13は、その分類に属する総ての項目を大きく画面表示し、タッチ入力部1からの項目指定入力を受け付けるようにする。



【特許請求の範囲】

【請求項1】利用者からの音声による選択すべき項目の入力等に用いられる音声入力手段と、この音声入力手段により入力された音声を認識する音声認識手段と、各種選択項目からなるメニュー及び各種メッセージ等の表示に用いられる表示手段と、この表示手段に表示されている画面上の項目を利用者の指等で位置指示することにより当該項目を指定入力するためのタッチ入力手段と、前記音声入力手段、前記音声認識手段、前記表示手段及び前記タッチ入力手段を制御して、前記音声入力手段及び前記タッチ入力手段からの入力の受け付けを行う制御手段とを具備し、

前記制御手段は、前記表示手段に前記選択項目からなる所定のメニューを表示した場合には、当該メニューの中で前記タッチ入力手段により指定入力可能な項目を限定し、前記タッチ入力手段により指定入力可能な項目の数を前記音声入力手段により入力可能な項目よりも少なく設定することを特徴とするマルチモーダル装置。

【請求項2】前記所定のメニュー中の項目は分類されて表示されており、

前記制御手段は、前記各分類を代表する項目については、前記音声入力手段及び前記タッチ入力手段のいずれからも入力可能として扱い、前記各分類に属する項目については、前記音声入力手段からのみ入力可能として扱うことを特徴とする請求項1記載のマルチモーダル装置。

【請求項3】前記所定のメニュー中には、前記分類に属する項目が1つだけで、それが当該分類を代表する項目となっているものが含まれており、

前記制御手段は、当該代表項目については、前記タッチ入力手段からも入力可能として扱うことを特徴とする請求項2記載のマルチモーダル装置。

【請求項4】前記制御手段は、前記所定のメニュー中の前記分類を代表する項目が、前記音声入力手段または前記タッチ入力手段から入力された場合には、当該分類に属する項目が前記所定のメニュー中の同じ項目より大きいサイズに設定された下位層のメニューを前記表示手段に表示し、この表示された下位層のメニュー中の項目を前記タッチ入力手段から入力可能として扱うことを特徴とする請求項3記載のマルチモーダル装置。

【請求項5】利用者からの音声による選択すべき項目の入力等に用いられる音声入力手段と、この音声入力手段により入力された音声を認識する音声認識手段と、各種選択項目からなるメニュー及び各種メッセージ等の表示に用いられる表示手段と、この表示手段に表示されている画面上の項目を利用者の指等で位置指示することにより当該項目を指定入力するためのタッチ入力手段と、前記音声入力手段、前記音声認識手段、前記表示手段及び前記タッチ入力手段を制御して、前記音声入力手段及び前記タッチ入力手段からの入力の受け付けを行う制御手段とを具備し、

段とを具備し、

前記制御手段は、前記音声入力手段によって入力可能な項目を前記表示手段に表示せずに、或いは一部のみを表示して、前記音声入力手段からの項目入力の受け付けを行い、入力音声に対して前記音声認識手段でのリジェクトが発生もしくは所定回数続いた場合、或いは認識結果に対する訂正要求が入力された場合には、前記タッチ入力手段による指定入力の対象となる項目を総て前記表示手段に表示して、前記タッチ入力手段からの項目入力の受け付けに切り換えることを特徴とするマルチモーダル装置。

【請求項6】利用者からの音声による選択すべき項目の入力等に用いられる音声入力手段と、この音声入力手段により入力された音声を認識する音声認識手段と、各種選択項目からなるメニュー及び各種メッセージ等の表示に用いられる表示手段と、この表示手段に表示されている画面上の項目を利用者の指等で位置指示することにより当該項目を指定入力するためのタッチ入力手段と、前記音声入力手段、前記音声認識手段、前記表示手段及び前記タッチ入力手段を制御して、前記音声入力手段及び前記タッチ入力手段からの入力の受け付けを行う制御手段とを具備し、

前記制御手段は、前記音声入力手段によって入力可能な項目を前記表示手段に表示せずに、或いは一部のみを表示して、前記音声入力手段からの項目入力の受け付けを行い、所定時間内に入力となされない場合には、前記タッチ入力手段による指定入力の対象となる項目を総て前記表示手段に表示して、前記タッチ入力手段からの項目入力の受け付けに切り換えることを特徴とするマルチモーダル装置。

【請求項7】利用者からの音声による選択すべき項目の入力等に用いられる音声入力手段と、この音声入力手段により入力された音声を認識する音声認識手段と、各種選択項目からなるメニュー及び各種メッセージ等の表示に用いられる表示手段と、この表示手段に表示されている画面上の項目を利用者の指等で位置指示することにより当該項目を指定入力するためのタッチ入力手段と、前記音声入力手段、前記音声認識手段、前記表示手段及び前記タッチ入力手段を制御して、前記音声入力手段及び前記タッチ入力手段からの入力の受け付けを行う制御手段とを具備し、

前記制御手段は、前記音声入力手段によって入力可能な項目を前記表示手段に表示せずに、或いは一部のみを表示して、前記音声入力手段からの項目入力の受け付けを行い、入力音声に対して前記音声認識手段でのリジェクトが発生もしくは所定回数続いた場合、認識結果に対する訂正要求が入力された場合、或いは所定時間内に入力となされない場合には、前記入力可能な項目を総て前記表示手段に表示して、前記音声入力手段からの項目入力の受け付けを継続することを特徴とするマルチモーダル

装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、利用者との対話手段を複数有し、複雑な操作をより簡単に利用可能とするマルチモーダル装置に関する。

【0002】

【従来の技術】近年、操作性の向上、或いは操作の自由度の拡大のために、音声入力（音声認識）手段及びタッチ入力手段などの複数の入力手段を備えたマルチモーダル装置が各欄開発されている。

【0003】この種のマルチモーダル装置では、選択入力の対象となる項目の一覧（メニュー）を画面表示して、その中から音声入力またはタッチ入力により任意の項目を入力させるのが一般的である。ここで、項目数が多い場合には、隣接項目（の表示領域）間の間隔が小さくなる。このような場合、タッチ入力では、視差の影響と指の太さの影響で正確なタッチが不可能となり、利用者の意図した項目の入力が行い難い。

【0004】そこで従来のマルチモーダル装置では、画面表示する項目数が多い場合には、多項目からの項目選択に当たらないタッチ入力手段からの入力を受け付けられないなどの制限を加え、1回に画面表示する項目数を少なくして、その項目については、音声入力及びタッチ入力の両入力を可能とするなどの方式が採られている。

【0005】しかし、前者の方式では、タッチ入力手段による項目入力は受け付けられないことから、利用者の操作の自由度に乏しく、また項目が画面表示されているために利用者が誤ってタッチ入力操作を行う可能性があり、しかもその操作は無駄になるという問題があった。一方、後者の方式では、対象となる項目数に影響されずに項目入力が行えるという音声入力手段の特長を十分に生かせないという問題があった。

【0006】また、音声入力では、利用者によっては音声認識でのリジェクトや認識誤りが発生し易く、このような利用者は、その後どうしたらよいか迷ってしまうが、従来のマルチモーダル装置では、そのような利用者に対する配慮はなされていなかった。

【0007】

【発明が解決しようとする課題】上記したように、従来のマルチモーダル装置では、複数の入力手段の連携が必ずしもうまくとれる構成となっておらず、有効に機能していなかった。即ち従来のマルチモーダル装置では、複数の入力手段が存在するがために、利用者にとっては、どの入力手段を用いて、どのように操作してよいのか分からなかったり、利用者に対する操作説明に時間がかかったり、異なる入力手段で同様の項目を入力させるために互いに入力機能の制限を加える結果となることも多く、必ずしも使い易いものではなかった。

【0008】本発明は上記事情を考慮してなされたものでその目的は、異なる入力手段である音声入力手段とタッチ入力手段との連携を各入力手段の特長を生かすようにとることができ、もって利用者の入力操作が容易且つ確実に行える使い勝手の良いマルチモーダル装置を提供することにある。

【0009】

【課題を解決するための手段及び作用】本発明のマルチモーダル装置は、表示手段に表示した複数の選択項目からメニューが所定のメニュー（予め定められた特定のメニュー）の場合には、当該メニューの中で前記タッチ入力手段により指定入力可能な項目を限定し、タッチ入力手段により指定入力可能な項目の数を音声入力手段により入力可能な項目よりも少なく設定することを特徴とするものである。

【0010】このような構成においては、所定メニュー上に多数の項目が含まれていても、視差の影響と指の太さの影響で正確なタッチ入力が不可能となるような表示領域の小さい項目についてのみ、タッチ入力手段による指定入力の対象外として扱うことにより、他の表示領域の大きい項目についてはタッチ入力手段により正しく指定入力することが可能となる。また、音声入力手段は、タッチ入力手段に対してなされた入力可能項目の制限の影響を受けず、したがって多項目からの項目選択に過しているという音声入力手段の特長も生かせる。

【0011】また本発明は、上記所定メニュー中の項目を分類して表示し、各分類を代表する項目については、音声入力手段及びタッチ入力手段のいずれからも入力可能として扱い、各分類に属する項目については、音声入力手段からのみ入力可能として扱うことも特徴とし、更に分類を代表する項目が、音声入力手段またはタッチ入力手段から入力された場合には、当該分類に属する項目が上記所定メニュー中の同じ項目より大きいサイズに設定された下位層のメニューを表示して、この表示された下位層のメニュー中の項目をタッチ入力手段から入力可能として扱うことも特徴とする。

【0012】このようにすることにより、分類を代表する項目については、初めからタッチ入力手段による入力も行え、また、その代表項目が音声入力またはタッチ入力されたならば、対応する分類に属する項目をタッチ入力手段から入力することも可能となる。

【0013】また本発明は、上記所定メニュー中に、分類に属する項目が1つだけで、それが当該分類を代表する項目となっているものを含め、この項目については、タッチ入力手段からも入力可能として扱うことも特徴とする。

【0014】このような構成においては、選択される頻度の高い項目を、上記所定メニュー上で分類に属する唯一の項目として扱うことにより、当該項目をタッチ入力によっても1回で選択することが可能となる。

【0015】また本発明は、音声入力手段によって入力可能な項目を表示せずに、或いは一部のみを表示して、音声入力手段からの項目入力の受け付けを行い、入力音声に対して音声認識処理でリジェクトが発生もしくは所定回数続いた場合、認識結果に対する訂正要求が入力された場合、或いは所定時間内に入力となされない場合のいずれかに該当する場合に、タッチ入力手段による指定入力の対象となる項目を総て表示して、タッチ入力手段からの項目入力の受け付けに切り換えることを特徴とする。

【0016】このような構成によれば、入力音声が正しく音声認識され難い傾向にある利用者、或いは音声の入力操作に迷って音声入力動作を起こさない利用者にとっては、自動的にタッチ入力が可能な状態に切り換えられるため、直ちにタッチ入力操作を行うことが可能となる。また、音声入力手段による入力が可能な項目の総てが表示されるのではないため、本発明が駅の券売機に適用される場合のように、入力すべき項目（駅名）を利用者が予め分かっているような場合には、その利用者に対し、総ての項目を表示する場合に比べて煩雑な印象を与えることが防止できる。

【0017】また、入力可能な項目を総て表示した場合に、上記のようにタッチ入力手段からの項目入力の受け付けに切り換えなくて、音声入力手段からの項目入力の受け付けを継続するようにしても良い。このような構成では、入力すべき項目を利用者が確実に把握していないような場合に、利用者の音声入力の手助けとすることができ。

【0018】

【実施例】

【第1実施例】以下、本発明の第1実施例を、ハンバーガショップの注文機に用いるマルチモーダル装置に適用した場合を例に、図面を参照して説明する。

【0019】図1は同実施例に係るマルチモーダル装置の概略構成を示すブロック図である。図1に示すマルチモーダル装置は、音声入力部11、表示部12、制御部13、音声認識部14、タッチ入力部15、音声出力部16及び接近検知部17から構成される。

【0020】音声入力部11は、利用者が発声した音声を入力するために用いられるもので、例えばマイクロホンが内蔵されたハンドセット型音声入力装置などである。表示部12は、各種選択項目（入力項目）からなるメニュー及び各種メッセージ等の表示に用いられるもので、例えばCRTディスプレイ、或いは液晶ディスプレイなどである。

【0021】制御部13は、表示部12及び音声出力部16を通して利用者との対話を管理・制御するものである。制御部13は、接近検知部17の検知結果に従う表示部12への初期画面の表示、音声認識部14の音声認識による入力並びにタッチ入力部15からのタッチ入力

の受け付け制御、受け付けた入力内容（と現在の状態）に従う表示画面の切り換え、及び表示部12並びに音声出力部16へのメッセージ出力等を司る。

【0022】音声認識部14は、音声入力部11から入力された音声の認識処理を司る。タッチ入力部15は、例えば表示部12の表示面に重ねて用いられるタッチパネルを有している。タッチ入力部15は、利用者が表示部12上の表示画面を見て所望の項目の表示位置を表示面に重ねられたタッチパネル上でタッチする（指等で触れる）ことにより、そのタッチ位置を辞書容量の変化、赤外線遮断、重力の変化などを利用して検知して入力するものである。このタッチ入力部15により入力された位置情報に相当する項目が何であるかは、即ち利用者により選択（入力）された項目が何であるかは、その位置情報と表示中の画面の構成情報をもとに、制御部13により認識されるものであるが、ここでは説明を簡単にするために、タッチ入力部15により項目が指定入力されるものとする。

【0023】音声出力部16は、音声によるメッセージ等の出力に用いられるもので、例えばスピーカである。接近検知部17は、利用者が本装置に接近したことを検出するものであり、例えば光学的なセンサである。

【0024】次に図1の構成のマルチモーダル装置の利用手順の一例を、図2及び図3のフローチャートと図4乃至図6の表示画面例を参照して説明する。まず、利用者が本装置に近づくと（ステップ201）、その旨が接近検知部17により検知されて（ステップ202）、制御部13に通知される。

【0025】すると制御部13は、図4に示すような注文用の表示画面（初期品名表）を表示部12に表示する（ステップ203）。この図4に示す表示画面は、多数の注文項目を符号401～406に示す6つの分類に分けて表示するための画面構造を有している。図4の表示画面例では、分類401～406を代表する項目（分類項目）として、「ハンバーガ」、「ジュース」、「スナック」、「クールドリンク」、「カツカレー」、「ホットドリンク」が用いられている。画面下方には、利用者の注文内容（に対する認識結果）を表示するための領域（＜ご注文＞欄）が確保されている。図では、注文内容として「チーズバーガ 1個」が示されているが、実際には、利用者が注文をし、その内容が認識された際に表示されるものである。また、この領域の右端には、注文の終了をタッチ入力により指示するための「確認」ボタン407が表示されている。なお、品名（項目）を示す文字列の他に、対応する品物の特徴を表した絵（マーク）を添えて、その品物の内容を分かり易くすることが可能である。

【0026】図4に示した表示画面では、利用者は音声入力部11を用いて音声による項目入力で注文すること、画面中の所望の項目の表示領域に相当するタッチ入

力部15（のタッチパネル上）の位置（以下、単に項目の領域と呼ぶ）を指で触れて（タッチして）注文することも可能である。

【0027】但し、音声による項目入力であれば、分類401～406に属する項目（品名）を1回の発声で特定できるものの、タッチ入力部15を用いた場合には、最初は分類項目しか指定入力できず、1回のタッチ入力では、所望の項目（品名）を特定できないようになってい

る。【0028】このように、1回のタッチ入力で目的の項目を特定できない理由は、表示部12として通常の小画面のディスプレイを用いた場合には、図4の例のように項目数が多いと、項目（の表示領域）のサイズ及び隣接項目（の表示領域）間の間隔が小さくなり、視差の影響と指の太さの影響で正確なタッチが不可能となることによる。このため本実施例では、分類項目をタッチした場合には（後述するような特別の分類項目の場合を除き）、その分類のみを大きく画面表示して、その分類に属する項目（品名）をタッチ入力できるようにしている。

【0029】具体的には、利用者は、音声入力部11に内蔵のマイクロホンに例えば「チーズバーガ」「1個」と発声して（ステップ204）注文しても良いし、タッチ入力部15上で「チーズバーガ」等を代表する分類項目である「ハンバーガ」の領域（分類401の領域）を触れる（ステップ209）ようにしても良い。

【0030】利用者が発声した音声は音声入力部11により入力される。音声認識部14は、音声入力部11により入力された音声の認識処理を行い、その認識結果を制御部13に渡す。

【0031】一方、利用者がタッチした項目の位置情報はタッチ入力部15により入力されて、制御部13に渡される。制御部13は、この位置情報と表示中の画面の構成情報をもとに、利用者が入力指定（選択）した項目を認識する。

【0032】さて上記の例において、音声による注文では、それが音声認識部14にて正しく認識されるならば、「チーズバーガ」を1回の発声で特定できる。これに対して、タッチ入力による注文では、前記した理由で1回のタッチでは特定できない。そこで制御部13は、上記の例のように「ハンバーガ」という分類401の領域がタッチされた場合には、図5に示すような「ハンバーガ」の分類に属する項目がタッチ入力可能なように拡大された表示画面を、表示部12に表示する（ステップ210）。

【0033】この図5に示す表示画面は、符号501に示すように、「ハンバーガ」の分類を大きく表示するための画面構造を有している。また当該画面には、注文品の個数をタッチ入力により指定するための個数入力領域502が確保されている。この個数入力領域502に

は、注文の取消を指示するための「取消」ボタン503が設けられている。また、画面下方には、利用者の注文内容の表示領域が確保されている。図では、注文内容として「チーズバーガ 1個」が示されているが、実際には、利用者が注文をし、その内容が認識された際に表示されるものである。また、この領域の右端には、注文の追加をタッチ入力により指示するための「追加」ボタン504と、注文の終了をタッチ入力により指示するための「確認」ボタン505が表示されている。

【0034】利用者は、図5に示す表示画面が表示されると、「ハンバーガ」の分類501中の所望の項目（品名）、例えば「チーズバーガ」をタッチし、続いて個数入力領域502上で、例えば「1」「個」とタッチすることによって（ステップ211）、タッチ入力部15から注文を行うことができる。

【0035】制御部13は、このタッチ入力結果から、利用者の注文した品と個数を認識する。ここで、例えばジュースを追加注文したい場合には、利用者はまず、「追加」ボタン504をタッチする（ステップ301）。すると制御部13は、再び図4に示した表示画面を表示する（ステップ302）。

【0036】次にこの状態で、利用者が、「オレンジジュース」等を代表する分類項目である「ジュース」の領域（分類402の領域）をタッチすると（ステップ303）、制御部13は、「ジュース」の分類に属する項目がタッチ入力可能なように拡大された（図5と同様の構造の）表示画面を表示部12に表示する（ステップ304）。

【0037】ここで利用者が、希望の品名と個数とを、例えば「オレンジジュース」「1」「個」の各領域をタッチすることによって指定入力すると（ステップ305）、制御部13は、「オレンジジュース 1個」が追加注文されたことを認識する。

【0038】また、「オレンジジュース 1個」を追加した状態で、先に注文した「チーズバーガ」を「取消」したい場合は、利用者はまず、図5の画面上の「追加」ボタン504をタッチした場合（ステップ301）と同様に、「ジュース」の分類の表示画面上の「追加」ボタンをタッチする（ステップ306）。すると、制御部13により、図4の表示画面が再表示される（ステップ307）。

【0039】ここで利用者が、「ハンバーガ」という分類401の領域をタッチすると（ステップ308）、図5の表示画面が再表示される（ステップ309）。利用者が、この状態で「チーズバーガ」「取消」とタッチすると（ステップ310）、制御部13は、先に受け付けた「チーズバーガ 1個」の注文の取消を行う。

【0040】そして、利用者が注文を終えたと判断したならば、「確認」ボタン505をタッチする（ステップ311）。これにより制御部13は、利用者からの注

文受け付けを終了する。

【0041】なお、図5中の<ご注文>欄の注文内容「チーズバーガ 1個」を1つの入力項目として扱うようにし、この「チーズバーガ 1個」の領域と、「取消」ボタン503とを連続してタッチすることにより、注文済みの「チーズバーガ」の取消を行うようにしても良い。この場合、利用者による取消の操作手順が簡略化される。

【0042】一方、音声入力の場合の「オレンジジュース」の追加注文であれば、(タッチ入力による追加注文と異なって)利用者は単に「オレンジジュース」「1個」と発声する(ステップ205)だけで良い。また、先に注文した「チーズバーガ」を取消する場合は、例えば「チーズバーガ」「取消」と発声する(ステップ206)だけで良い。ここで注文の終了指示は、「確認」と発声しても(ステップ207)、「確認」ボタンをタッチしても(ステップ208)行える。

【0043】なお、図4の例では、多数の注文項目を6つの分類401～406に分けて表示しているが、「カツカレー」のように、1つの品名(項目)をそのまま分類項目(分類を代表する項目)とした場合には、他の「ハンバーガ」などの分類項目と異なり、その分類項目「カツカレー」(の領域)をタッチするだけで、1つに特定することが可能となる。

【0044】また、図4に示した画面の表示状態において、分類項目のいずれか1つ、例えば「ジュース」を発声した場合に、ジュースの分類に属する項目だけを、図6に示すように大きく表示させるようにすることもできる。この図6に示す画面の表示状態で、例えば「オレンジジュース 1個」を注文するには、「オレンジジュース」「1個」と発声しても、「オレンジジュース」「1」「個」とタッチしても良い。

【第2実施例】次に、本発明の第2実施例を、駅の券売機に用いるマルチモーダル装置に適用した場合を例に、図7のフローチャートと図8乃至図12の表示画面例を参照して説明する。ここで、本装置の概略構成は、図1に示した前記第1実施例に係る装置と同様であるため、第2実施例の説明においては、便宜上図1も併用する。

【0045】まず、利用者が本装置に近づき(ステップ701)、その旨が接近検知部17により検知されると(ステップ702)、制御部13により表示部12に図8に示すような表示画面(案内画面)が表示される(ステップ703)。この図8に示す画面では、個々の駅名が表示されていない。その理由は、第1実施例で述べたようなハンバーガショップの注文機と違って、利用者は、一般に自分の行きたい駅の駅名を知っているからであり、表示の必要がないためである。また、逆に個々の駅名を表示したのでは、かえって利用者に煩雑な印象を与えることが多いためである。

【0046】利用者は、図5に示す案内画面が表示され

ると、その案内「どちらまででしょうか。」に従って、音声入力部11に内蔵のマイクロホンに、例えば「浅草橋」「1枚」(利用者が子供の場合には、「浅草橋」「子供」「1枚」)というように発声する(ステップ704)。

【0047】利用者が発声した音声は音声入力部11により入力される。音声認識部14は、音声入力部11により入力された音声「浅草橋」「1枚」の認識処理を行い(ステップ705)、その認識結果を制御部13に渡す。すると制御部13は、その音声認識結果を受けて、例えば図9に示すように「浅草橋 大人 1枚」を含む認識結果メッセージを画面表示する(ステップ706)。

【0048】利用者は、図9に示す画面を参照し、行先(及び枚数)が正しく認識されたかと判断できたなら、画面中に設けられた「確認」ボタン902をタッチする(ステップ707)。すると制御部13は、料金を計算し、利用者に対してその料金の支払いを促すメッセージを表示する。

【0049】一方、正しい認識結果が表示されていなかった場合には、利用者は、図9の画面中に設けられた「訂正」ボタン901をタッチする(ステップ708)。すると制御部13は、図8に示した案内画面を再表示する(ステップ709)。但し、「いちっしょいませ。」のメッセージ部分は消す。

【0050】ここで、利用者が、再び「浅草橋」「1枚」と発声しても、即ち再度音声入力を行っても、図9に示したような正しい認識結果が表示されなかった場合には、利用者は、再び「訂正」ボタン901をタッチする(ステップ710)。

【0051】制御部13は、訂正回数をカウントしており、所定回数、例えば本実施例のように2回「訂正」がタッチ入力されると、現在の利用者では、これ以上音声による項目入力(ここでは、行先、枚数などの入力)を繰り返しても、正しく認識できないものと判断し、図10に示すような表示画面に切り換えて、タッチでの項目入力の受け付けに切り換える(ステップ711)。この図10の表示画面は、駅名の頭文字により、駅名のタッチ入力可能な画面構造となっている。

【0052】なお、利用者が所定回数音声入力しても、音声認識部14にてリジェクトされる場合(即ち音声認識結果を出せない場合)、或いは1つに決められない場合にも、上記した「訂正」回数が所定回数(2回)となった場合と同様の扱いをしてタッチ入力に切り換えるようにしても良い。

【0053】また、図8の案内画面を表示した後、或る一定時間を経過しても、(利用者が行先の駅名をはっきりと覚えていないとか、音声入力の仕方が分からないなどの理由で)音声入力がなかった場合に、上記した「訂正」回数が所定回数となった場合と同様の扱いをしてタ

タッチ入力に切り換えるようにしても良い。勿論、1度目は再度案内画面を表示して、利用者に注意を喚起した後、タッチ入力に切り換えるようにしても良い。

【0054】さて利用者は、図10に示す画面が表示されると、希望する駅名「浅草橋」の頭文字「あ」をタッチ入力する(ステップ712)。すると制御部13は、図11に示すような、頭文字が「あ」の駅名を入力項目とするタッチ入力用の一覧を表示する(ステップ713)。

【0055】そこで利用者は、「浅草橋」をタッチ入力する(ステップ714)。図11の一覧画面には、図5中の個数入力領域502と同様な枚数入力領域(図では省略)が確保されており、この領域上で、切符の枚数、頁には「子供(或いは大人/子供)のタッチ入力が可能となっている。

【0056】利用者によるタッチ入力結果は、図9と同様の画面に表示され、利用者の希望通りの結果であるならば、利用者は「確認」ボタンをタッチする。この「確認」ボタンのタッチにより、「訂正」回数のカウント結果がクリアされる。また、この「訂正」回数のカウント結果は、上記したタッチ入力への切り換えによりクリアされるようにしても良い。

【0057】なお、上記のようにタッチ入力に切り換える代わりに、図12に示すような、総ての駅名の一覧を画面表示(一度に表示し切れない場合には、次ページ指定により画面を切り換えて表示)し、利用者が音声で入力する手助とするようにしても良い。この方式は、利用者が行先の駅名をはっきりと覚えていないような場合に適する。

【0058】

【発明の効果】以上詳述したようにこの発明によれば、多数の項目の中から所望の項目を選択して入力するには適しているが、利用者によっては入力内容が正しく認識され難い音声入力手段と、多数の項目の中からの項目の選択入力には適していないが、少数の項目の中から選択

*する際には利用者によらずに、簡単かつ確実に入力できるタッチ入力手段の特長を生かすように、これら両入力手段の接続をとることができるため、利用者の入力操作が容易且つ確実に行え、使い勝手を向上することができる。

【図面の簡単な説明】

【図1】本発明の第1実施例に係るマルチモーダル装置の概略構成を示すブロック図。

【図2】同実施例における利用手順の一例を説明するためのフローチャートの一部を示す図。

【図3】同実施例における利用手順の一例を説明するためのフローチャートの残りを示す図。

【図4】同実施例において多数の項目を分類して表示する際の画面表示例を示す図。

【図5】同実施例においてタッチ入力により選択された1つの分類に属する項目を大きく表示する際の表示画面例を示す図。

【図6】同実施例において音声入力により選択された1つの分類に属する項目を大きく表示する際の表示画面例を示す図。

【図7】本発明の第2実施例における利用手順の一例を説明するためのフローチャート。

【図8】同実施例における案内画面の一例を示す図。

【図9】同実施例における認識結果表示画面の一例を示す図。

【図10】同実施例におけるタッチ入力用画面の一例を示す図。

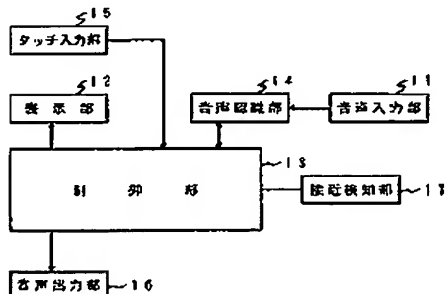
【図11】図10の画面の下位層をなすタッチ入力用画面の一例を示す図。

【図12】同実施例における音声入力用画面の一例を示す図。

【符号の説明】

11…音声入力部、12…表示部、13…制御部、14…音声認識部、15…タッチ入力部、16…音声出力部、17…接近検知部。

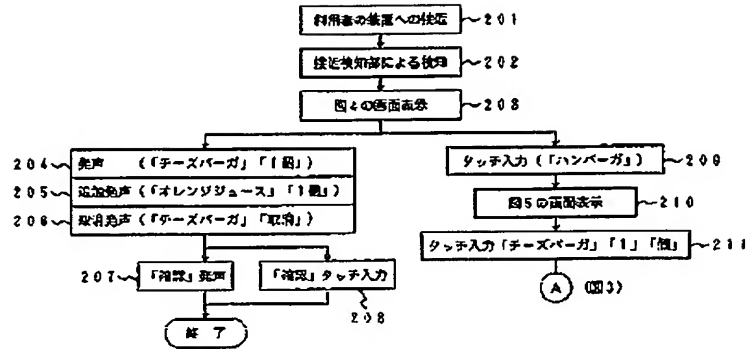
【図1】



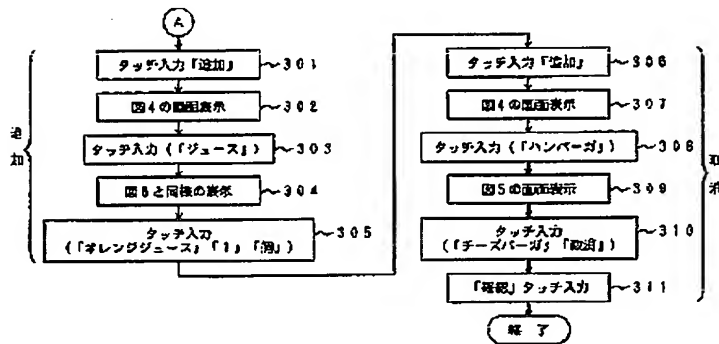
【図8】



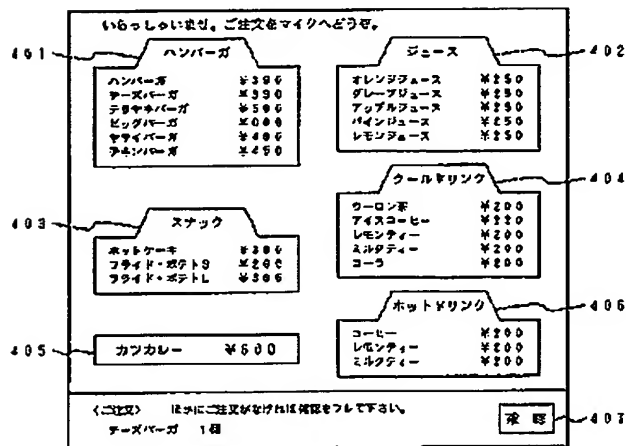
【図2】



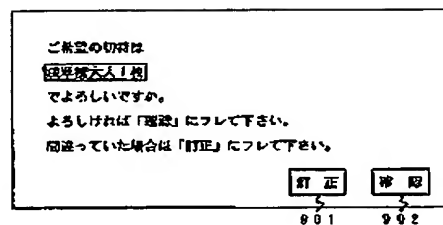
【図3】



【図4】



【図9】



(9)

特開平7-110751

【図5】

品名をフレたあと、値段をフレして下さい。

ハンバーグ

品名	値段
ハンバーグ	¥390
チーズバーグ	¥390
テリヤキバーグ	¥500
ビッグバーグ	¥600
ウサイバーグ	¥400
チキンバーグ	¥450

1	2	3
4	5	6
7	8	9
0	取消	戻

501

502

503

504

505

＜戻す＞ すでに注文された品は消滅をフレして下さい。
チーズバーグ 1個

追加 削除

【図10】

駅名の頭文字をフレして下さい。

あ	か	さ	た	な	は	ま	や	ら	わ
い	リ	
う	る	
え	れ	
お	る	

【図6】

『オレンジジュース』『1個』のようにマイクへどうぞ。

ジュース

品名	値段
オレンジジュース	¥250
グレープジュース	¥250
アップルジュース	¥250
パイナップルジュース	¥250
レモンジュース	¥250

＜戻す＞

【図11】

希望の駅名をフレして下さい。

相 原	上 尾
青 嶺	阿 佐ヶ 谷
赤 塚	浅 草 橋
赤 羽	旗 本
秋 川	旭
昭 島	足 尾
狹 野 路	次 々 々

【図12】

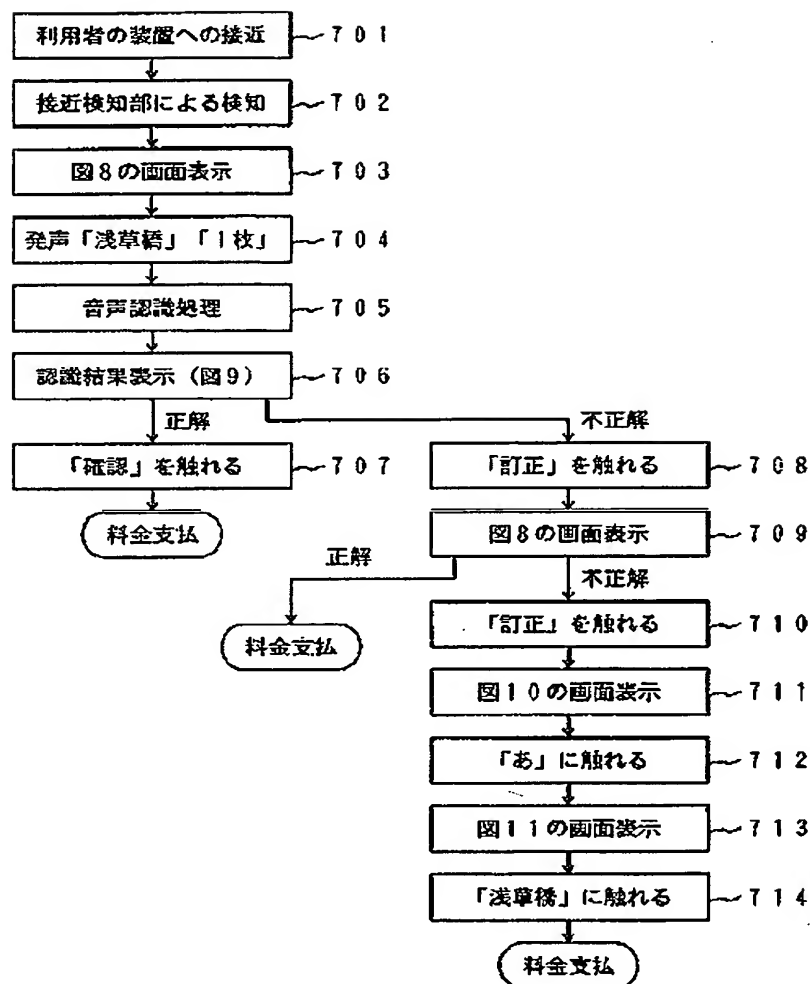
希望の駅名をマイクへどうぞ。

あ	相 原	580
い	青 嶺	140
う	赤 塚	120

(10)

特開平 7-110751

【図 7】



PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-110751

(43)Date of publication of application : 25.04.1995

(51)Int.Cl. G06F 3/14
G06F 3/16

(21)Application number : 05-254218 (71)Applicant : TOSHIBA CORP

(22)Date of filing : 12.10.1993 (72)Inventor : MATSUURA
HIROSHI
KAMIO HIROYUKI

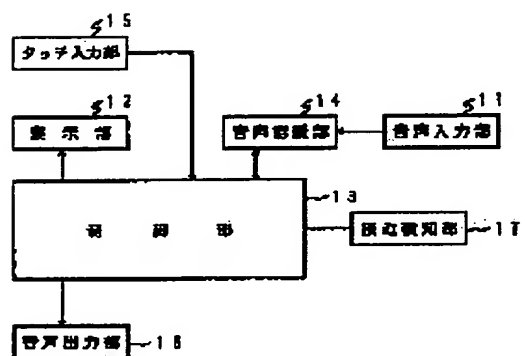
(54) MULTIMODAL DEVICE

(57)Abstract:

PURPOSE: To easily and surely perform the input operation by a user and to improve operability by linking a voice input means with a touch input means.

CONSTITUTION: When the approach of the user is detected by an approach detection part 17, a control part 13 classifies and displays a lot of input items on a display part 12 and receives the input of all items by sound from a sound input part 11, however, receives only the representative items of classification by touch input from a touch input part 15.

When one of the representative items is touch inputted, the control part 13 enlarges and displays all the items belonging to the classification on a screen and receives item specifying input from the touch input part 15.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's

decision of rejection]

[Kind of final disposal of application
other than the examiner's decision
of rejection or application
converted registration]

[Date of final disposal for
application]

[Patent number]

[Date of registration]

[Number of appeal against
examiner's decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and INPIT are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation
may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The voice input means used for the input of the item with the
voice from a user which should be chosen etc., A speech recognition
means to recognize the voice inputted by this voice input means, The
display means used for the display of the menu which consists of
various selections, various messages, etc., The touch input means for
carrying out the assignment input of the item concerned by carrying out
the location directions of the item on the screen currently displayed on
this display means with a user's finger etc., Said voice input means, said
speech recognition means, said display means, and said touch input
means are controlled. The control means which receives the input from
said voice input means and said touch input means is provided. Said
control means When the predetermined menu which becomes said

display means from said selections is displayed Multi-modal equipment characterized by limiting the item in which an assignment input is possible with said touch input means in the menu concerned, and setting up the number of the items in which an assignment input is possible with said touch input means fewer than the item which can be inputted with said voice input means.

[Claim 2] It is multi-modal equipment according to claim 1 characterized by classifying and displaying the item in said predetermined menu, and for either said voice input means and said touch input means treating said control means as an input being possible about the item representing said each classification, and treating as an input being possible only from said voice input means about the item belonging to said each classification.

[Claim 3] It is multi-modal equipment according to claim 2 which that from which the item belonging to said classification is the item in which it represents the classification concerned only with one is contained in said predetermined menu, and is characterized by treating said control means as an input being possible also from said touch input means about the representation item concerned.

[Claim 4] When the item representing said classification in said predetermined menu is inputted from said voice input means or said touch input means, said control means The menu of a lower layer with which the item belonging to the classification concerned was set as larger size than the same item in said predetermined menu is displayed on said display means. Multi-modal equipment according to claim 3 characterized by treating the item in the menu of this displayed lower layer as an input being possible from said touch input means.

[Claim 5] The voice input means used for the input of the item with the voice from a user which should be chosen etc., A speech recognition means to recognize the voice inputted by this voice input means, The display means used for the display of the menu which consists of various selections, various messages, etc., The touch input means for carrying out the assignment input of the item concerned by carrying out the location directions of the item on the screen currently displayed on this display means with a user's finger etc., Said voice input means, said speech recognition means, said display means, and said touch input means are controlled. The control means which receives the input from said voice input means and said touch input means is provided. Said control means Only a part is displayed without displaying the item which can be inputted with said voice input means on said display means. When

the item input from said voice input means was received and rejection with said speech recognition means continues generating or the number of predetermined times to input voice, Or multi-modal equipment characterized by displaying all the items set as the object of the assignment input by said touch input means on said display means, and switching to registration of the item input from said touch input means when the correction demand to a recognition result is inputted.

[Claim 6] The voice input means used for the input of the item with the voice from a user which should be chosen etc., A speech recognition means to recognize the voice inputted by this voice input means, The display means used for the display of the menu which consists of various selections, various messages, etc., The touch input means for carrying out the assignment input of the item concerned by carrying out the location directions of the item on the screen currently displayed on this display means with a user's finger etc., Said voice input means, said speech recognition means, said display means, and said touch input means are controlled. The control means which receives the input from said voice input means and said touch input means is provided. Said control means Only a part is displayed without displaying the item which can be inputted with said voice input means on said display means. When the item input from said voice input means is received and an input is not made in predetermined time Multi-modal equipment characterized by displaying all the items set as the object of the assignment input by said touch input means on said display means, and switching to registration of the item input from said touch input means.

[Claim 7] The voice input means used for the input of the item with the voice from a user which should be chosen etc., A speech recognition means to recognize the voice inputted by this voice input means, The display means used for the display of the menu which consists of various selections, various messages, etc., The touch input means for carrying out the assignment input of the item concerned by carrying out the location directions of the item on the screen currently displayed on this display means with a user's finger etc., Said voice input means, said speech recognition means, said display means, and said touch input means are controlled. The control means which receives the input from said voice input means and said touch input means is provided. Said control means Only a part is displayed without displaying the item which can be inputted with said voice input means on said display means. When the item input from said voice input means was received and rejection with said speech recognition means continues generating or the number

of predetermined times to input voice, when the correction demand to a recognition result is inputted, an input should do in predetermined time – the multi-modal equipment characterized by displaying all the items in which said input is possible on said display means, and continuing registration of the item input from said voice input means when there is nothing.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention has two or more dialogue means with a user, and relates to the multi-modal equipment which makes complicated actuation available more simply.

[0002]

[Description of the Prior Art] In recent years, various development of the multi-modal equipment equipped with two or more input means, such as a voice input (speech recognition) means and a touch input means, for amplification of the improvement in operability or the degree of freedom of actuation is carried out.

[0003] It is common to carry out a screen display of the list of the items set as the object of a selection input (menu) with this kind of multi-modal equipment, and to make the item of arbitration input by voice input or the touch input from that inside. Here, when there are many items, spacing between contiguity items (viewing area) becomes small. In such a case, in a touch input, an exact touch becomes impossible under the effect of parallax, and the effect of the size of a finger, and it is hard to perform the input of the item which the user meant.

[0004] Then, with conventional multi-modal equipment, when there were many items which carry out a screen display, the number of items which adds a limit of not receiving the input from a touch input means by which it is not suitable for item selection from many items, or carries out a screen display at once was lessened, and a method, such as enabling both the inputs of voice input and a touch input, was taken about the item.

[0005] However, by the former method, by being, since [for which the item input by the touch input means is not received] it was deficient in the degree of freedom of actuation of a user and a screen display of the item was carried out from things, it might perform touch alter operation accidentally [user], and moreover, the actuation had the problem of becoming useless. On the other hand, by the latter method, there was a problem that the features of a voice input means by which an item input can be performed without being influenced by the target number of items could not fully be employed efficiently.

[0006] Moreover, although it will waver [what such a user should do after that by being easy to generate rejection by speech recognition, and a recognition error for some users and] by voice input, the consideration to such a user was not made with conventional multi-modal equipment.

[0007]

[Problem(s) to be Solved by the Invention] As described above, with conventional multi-modal equipment, cooperation of two or more input means did not necessarily become the configuration which can be taken well, and was not functioning effectively. that is, it was not what results in adding a limit of an input function mutually in many cases in order are unclear in how may I use any input means for a user and I may operate it, or the operating instruction to a user takes time amount or to make the same item input with a different input means hard [slight / in which two or more input means exist with conventional multi-modal equipment], and it is not necessarily easy using .

[0008] This invention was made in consideration of the above-mentioned situation, and the object is in offering multi-modal equipment with the sufficient user-friendliness which can take cooperation with the voice input means and touch input means which are a different input means so that the features of each input means may be employed efficiently, has it, and can ensure [easily / a user's alter operation / and] it.

[0009]

[Means for Solving the Problem and its Function] two or more selections which displayed the multi-modal equipment of this invention on a display means -- when a menu [from] is a predetermined menu (the specific menu defined beforehand), the item in which an assignment input is possible limits with said touch input means in the menu concerned, and it is characterized by to set up the number of the items in which an assignment input is possible with a touch input means fewer than the item which can input with a voice-input means.

[0010] In such a configuration, even if many items are included on the predetermined menu, about the large item of other viewing areas, it becomes that it is possible to carry out an assignment input correctly with a touch input means by treating as outside of the object of the assignment input by the touch input means only about a small item of a viewing area it becomes impossible under the effect of parallax, and the effect of the size of a finger exact to touch input. Moreover, a voice input means is not influenced by the item which was made to the touch input means and which can be inputted of a limit, therefore can also employ efficiently the features of a voice input means by which it is suitable for item selection from many items.

[0011] moreover, about the item which this invention classifies and displays the item in the above-mentioned predetermined menu, and represents each classification About the item to which either a voice input means and a touch input means are treated as an input being possible, and they belong to each classification When the item which is characterized also by treating as an input being possible only from a voice input means, and represents a classification further is inputted from a voice input means or a touch input means The item belonging to the classification concerned is characterized also by displaying the menu of the lower layer set as larger size than the same item in the above-mentioned predetermined menu, and treating the item in the menu of this displayed lower layer as an input being possible from a touch input means.

[0012] By doing in this way, it also becomes possible about the item representing a classification to input the item to which the input by the touch input means can also be performed from the start, and the representation item belongs to voice input or the classification which corresponds if a touch input is carried out from a touch input means.

[0013] Moreover, this invention is characterized also by treating as an input being possible also from a touch input means about this item in the above-mentioned predetermined menu including the thing used as the

item with which the number of the items belonging to a classification is one, and it represents the classification concerned.

[0014] In such a configuration, it becomes possible to choose the item concerned at once also by the touch input by treating an item with the high frequency chosen as an only item which belongs to a classification on the above-mentioned predetermined menu.

[0015] Moreover, this invention displays only a part, without displaying the item which can be inputted with a voice input means. When the item input from a voice input means was received and rejection continues generating or the number of predetermined times by speech recognition processing to input voice, when the correction demand to a recognition result is inputted, an input should do in predetermined time -- in corresponding to either in case there is nothing, all the items set as the object of the assignment input by the touch input means are displayed, and it carries out ***** switched to registration of the item input from a touch input means as the description.

[0016] According to such a configuration, for the user in the inclination for the speech recognition of the input voice to be hard to be carried out correctly, or the user who wavers in audio alter operation and does not cause voice input actuation, since it is automatically switched to the condition in which a touch input is possible, it becomes possible to perform touch alter operation promptly. Moreover, since all the items in which the input by the voice input means is possible are not displayed, when the user understands beforehand the item (name of the station) which should be inputted like [in case this invention is applied to the ticket machine of a station], it can prevent giving a complicated impression to the user compared with the case where all items are displayed.

[0017] Moreover, you may make it continue registration of the item input from a voice input means without switching to registration of the item input from a touch input means as mentioned above, when all the items that can be inputted are displayed. With such a configuration, when the user does not grasp certainly the item which should be inputted, it can consider as the help of a user's voice input.

[0018]

[Example]

The case where the 1st example of this invention is applied to the multi-modal equipment used for the order machine of a hamburger shop is explained to an example with reference to a drawing below the [1st example].

[0019] Drawing 1 is the block diagram showing the outline configuration of the multi-modal equipment concerning this example. The multi-modal equipment shown in drawing 1 consists of the voice input section 11, a display 12, a control section 13, the speech recognition section 14, the touch input section 15, the voice output section 16, and the access detection section 17.

[0020] The voice input section 11 is the hand set mold audio input unit with which it is used in order to input the voice which the user uttered, and the microphone was built in. A display 12 is used for the display of the menu which consists of various selections (input item), various messages, etc., and is a CRT display or a liquid crystal display.

[0021] A control section 13 lets a display 12 and the voice output section 16 pass, and manages and controls a dialogue with a user. A control section 13 manages the message output to the voice output section 16 etc. at receptionist control of the touch input from the touch input section 15, a switch of the display screen according to the content of a reception beam input (current condition), and display 12 list in the input list by the display of the initial screen to the display 12 according to the detection result of the access detection section 17, and the speech recognition of the speech recognition section 14.

[0022] The speech recognition section 14 manages recognition processing of the voice inputted from the voice input section 11. The touch input section 15 has the touch panel used for the screen of a display 12 in piles. The touch input section 15 detects and inputs the touch location using change of electrostatic capacity, infrared electric shielding, change of gravity, etc. by what (a finger etc. describes) is touched on the touch panel with which the user looked at the display screen on a display 12, and put the display position of a desired item on the screen. Although recognized by the control section 13 based on that positional information and configuration information of a screen on display, in order for what the item equivalent to the positional information inputted by this touch input section 15 is, i.e., what is the item chosen by the user (input)?, to simplify explanation here, the assignment input of the item shall be carried out by the touch input section 15.

[0023] The voice output section 16 is used for outputs, such as a message with voice, and is a loudspeaker. The access detection section 17 detects that the user approached this equipment, and is an optical sensor.

[0024] Next, an example of the utilization procedure of the multi-modal

equipment of the configuration of drawing 1 is explained with reference to drawing 2 and the flow chart of drawing 3 and drawing 4 thru/or the example of the display screen of drawing 6 . First, if a user approaches this equipment (step 201), that will be detected by the access detection section 17 (step 202), and will be notified to a control section 13.

[0025] Then, a control section 13 displays the display screen for an order (initial name-of-article table) as shown in drawing 4 on a display 12 (step 203). The display screen shown in this drawing 4 has the screen structure for dividing and displaying many order items on six classifications shown in signs 401-406. In the example of a display screen of drawing 4 , a "hamburger", "juice", a "snack", a "cool drink", "cutlet curry", and a "hot drink" are used as an item (provisions of classification) representing classifications 401-406. The field (<order> column) for displaying the user's order content (receiving recognition result) is secured to the screen lower part. In drawing, it is "cheeseburger as an order content. Although 1 piece" is shown, actually, a user places an order, and when the content has been recognized, it is displayed. Moreover, the "check" carbon button 407 for directing termination of an order by the touch input is displayed on the right end of this field. In addition, it is possible to attach the picture (mark) showing the description of goods [/ other than the character string which shows a name of article (item)], and to make the content of the goods intelligible.

[0026] In the display screen shown in drawing 4 , a user can also touch and place [also placing an order in an item input with voice using the voice input section 11, and] an order for the location (it is only hereafter called the field of an item) of the touch input section 15 (on a touch panel) equivalent to the viewing area of the item of the request in a screen with a finger (touching).

[0027] However, if it is an item input with voice, although the item (name of article) belonging to classifications 401-406 can be specified by one utterance, when the touch input section 15 is used, only provisions of classification can carry out an assignment input at first, but a desired item (name of article) can be specified no longer in one touch input.

[0028] Thus, if the reason for the ability not to specify the target item in one touch input has many items like the example of drawing 4 when the display of the usual small screen is used as a display 12, it will be because the size of an item (viewing area) and spacing between contiguity items (viewing area) become small and the exact touch of them becomes impossible under the effect of parallax, and the effect of

the size of a finger. for this reason -- the case where provisions of classification are touched in this example -- (-- the case of special provisions of classification which are mentioned later -- removing --) -- a screen display only of that classification is carried out greatly, and it can be made to carry out the touch input of the item (name of article) belonging to that classification.

[0029] a user may utter with a "cheeseburger" and "one piece" to the microphone of built-in in the voice input section 11, and may place an order (step 204), and, specifically, the field (field of classification 401) of the "hamburger" which is the provisions of classification which represent a "cheeseburger" etc. on the touch input section 15 is touched (step 209) -- it is good even if like.

[0030] The voice which the user uttered is inputted by the voice input section 11. The speech recognition section 14 performs recognition processing of the voice inputted by the voice input section 11, and passes the recognition result to a control section 13.

[0031] On the other hand, the positional information of the item which the user touched is inputted by the touch input section 15, and is passed to a control section 13. A control section 13 recognizes the input assignment (selection) item [user] based on this positional information and the configuration information of a screen on display.

[0032] Now, in the above-mentioned example, by the order with voice, if it is correctly recognized in the speech recognition section 14, a "cheeseburger" can be specified by one utterance. On the other hand, at the order by the touch input, it cannot specify by one touch by the above mentioned reason. Then, a control section 13 displays the display screen expanded so that a touch input of the item belonging to a classification of a "hamburger" as shown in drawing 5 might be possible on a display 12, when touched in the field of the classification 401 of a "hamburger" like the above-mentioned example (step 210).

[0033] The display screen shown in this drawing 5 has the screen structure for displaying a classification of a "hamburger" greatly, as shown in a sign 501. Moreover, the number input area 502 for specifying the number of ordered goods by the touch input is secured in the screen concerned. "Cancellation" carbon button 503 for directing cancellation of an order is formed in this number input area 502. Moreover, the viewing area of the user's order content is secured to the screen lower part. In drawing, it is "cheeseburger as an order content. Although 1 piece" is shown, actually, a user places an order, and when the content has been recognized, it is displayed. Moreover, the "additional" carbon

button 504 for directing the addition of an order by the touch input and the "check" carbon button 505 for directing termination of an order by the touch input are displayed on the right end of this field.

[0034] A user can perform an order from (step 211) and the touch input section 15 by touching, and touching the item of the request under classification 501 of a "hamburger" (name of article), for example, a "cheeseburger", with "1" and an "individual" on the number input area 502 continuously, if the display screen shown in drawing 5 is displayed.

[0035] A control section 13 recognizes the elegance and the number which the user ordered from this touch input result. Here, a user touches the "additional" carbon button 504 first to order juice additionally (step S301). Then, a control section 13 displays the display screen again shown in drawing 4 (step 302).

[0036] Next, if the field (field of classification 402) of the "juice" which is the provisions of classification in which a user represents "orange juice" etc. with this condition is touched (step 303), a control section 13 will display the display screen (the same structure as drawing 5) expanded so that a touch input of the item belonging to a classification of "juice" might be possible on a display 12 (step 304).

[0037] When a user does the assignment input of the name of article and the number of hope by touching each field of "orange juice" "1" "individual" here (step 305), a control section 13 is "orange juice. It recognizes that 1 piece" was ordered additionally.

[0038] Moreover, "orange juice A user "touches [to cancel the "cheeseburger" ordered previously, where 1 piece" is added]" the "additional" carbon button on the display screen of a classification of "juice" first like the case (step 301) where the "additional" carbon button 504 on the screen of drawing 5 is touched (step 306). Then, regeneration of the display screen of drawing 4 is carried out by the control section 13 (step 307).

[0039] If a user touches the field of the classification 401 of a "hamburger" here (step 308), regeneration of the display screen of drawing 5 will be carried out (step 309). When a user touches with "cheeseburger" "cancellation" in this condition (step 310), a control section 13 is "cheeseburger received previously. The order of 1 piece" is canceled.

[0040] And if it judges that a user finished placing an order, the "check" carbon button 505 will be touched (step 311). Thereby, a control section 13 ends the order registration from a user.

[0041] in addition, the order content of the <order> column in drawing 5

— “cheeseburger one-piece” — as one input item — it can treat — making — this — “it may be made to cancel a “cheeseburger” by touching the field of cheeseburger one-piece”, and “cancellation” carbon button 503 continuously. [finishing / an order] In this case, the operating procedure of cancellation by the user is simplified.

[0042] on the other hand, if it is an additional order of the “orange juice” in the case of voice input, a user (differing from the additional order by the touch input) will only utter with “orange juice” and “one piece” — being sufficient (step 205) . moreover, when canceling the “cheeseburger” ordered previously, it utters with a “cheeseburger” and “cancellation” — being sufficient (step 206) . Termination directions of an order here can be performed even if it utters with “a check” (step 207), and it touches a “check” carbon button (step 208).

[0043] In addition, like “cutlet curry”, although many order items are divided and displayed on six classifications 401–406 in the example of drawing 4 , when one name of article (item) is made into provisions of classification (item representing a classification) as it was, unlike the provisions of classification of other “hamburgers” etc., it becomes possible to specify it as one only by touching the provisions of classification “cutlet curry” (field).

[0044] Moreover, in the display condition of the screen shown in drawing 4 , when uttered, any one, for example, “juice”, of provisions of classification, only the item belonging to a classification of juice can be greatly displayed, as shown in drawing 6 . It is in the display condition of the screen shown in this drawing 6 , for example, is “orange juice. In order to place an order for 1 piece”, it may utter with “orange juice” and “one piece”, or you may touch with “orange juice”, “1”, and an “individual.”

The case where the [2nd example], next the 2nd example of this invention are applied to the multi-modal equipment used for the ticket machine of a station is explained to an example with reference to the flow chart of drawing 7 , drawing 8 , or the example of the display screen of drawing 12 . Here, since the outline configuration of this equipment is the same as that of the equipment concerning said 1st example shown in drawing 1 , it also uses drawing 1 together for convenience in explanation of the 2nd example.

[0045] First, if a user approaches this equipment (step 701) and that is detected by the access detection section 17 (step 702), the display screen (initial screen format) as shown in a display 12 by the control section 13 at drawing 8 will be displayed (step 703). Each name of the

station is not expressed as the screen shown in this drawing 8 . It is because the user knows the name of the station to which he generally wants to go unlike the order machine of a hamburger shop which explained the reason in the 1st example, and is because there is no need for a display. Moreover, in having displayed each name of the station on reverse, it is for giving an impression complicated to a user on the contrary in many cases.

[0046] if the initial screen format which a user shows to drawing 5 is displayed -- the advice "to which is it?" -- following -- the microphone of built-in in the voice input section 11 -- for example, "Asakusabashi" and "one sheet" (the case where a user is a child -- "Asakusabashi", a "child", and "one sheet") -- as -- it utters (step 704).

[0047] The voice which the user uttered is inputted by the voice input section 11. The speech recognition section 14 performs recognition processing of the voice "Asakusabashi" inputted by the voice input section 11 and a "one sheet" (step 705), and passes the recognition result to a control section 13. Then, a control section 13 is a "Asakusabashi adult, in response to the fact that the speech recognition result, for example, as shown in drawing 9 . A screen display of the recognition result message containing 1 sheet" is carried out (step 706).

[0048] If a user is able to judge that the destination (and number of sheets) has been recognized correctly with reference to the screen shown in drawing 9 , he will touch the "check" carbon button 902 prepared all over the screen (step 707). Then, a control section 13 calculates a tariff and displays the message which demands payment of the tariff from a user.

[0049] On the other hand, when the right recognition result is not displayed, a user touches "correction" carbon button 901 prepared all over the screen of drawing 9 (step 708). Then, a control section 13 carries out regeneration of the initial screen format shown in drawing 8 (step 709). However, the message part of "welcome" is erased.

[0050] Here, when a right recognition result as shown in drawing 9 is not displayed whether a user utters with "Asakusabashi" and "one sheet" again or performs voice input again namely, a user touches "correction" carbon button 901 again (step 710).

[0051] By the present user, if the count of correction is counted and the touch input of the 2 times "correction" is carried out like the count of predetermined, for example, this example, even if a control section 13 repeats an item input (here input of a destination, number of sheets,

etc.) with voice more than this, it will judge it as what cannot be recognized correctly, will be switched to a display screen as shown in drawing 10 , and will be switched to registration of the item input in a touch (step 711). The display screen of this drawing 10 has screen structure in which the touch input of a name of the station is possible in the initial of a name of the station.

[0052] In addition, even if a user does count voice input of predetermined, when rejecting in the speech recognition section 14, or also when one is not decided on (namely, when a speech recognition result cannot be taken out), the same treatment as the case where the above-mentioned count of "correction" turns into a count of predetermined (2 times) is carried out, and you may make it switch to a touch input.

[0053] Moreover, after displaying the initial screen format of drawing 8 , even if it goes through a certain fixed time amount, the same treatment as the case where the above-mentioned count of "correction" turns into a count of predetermined when there is no voice input is carried out (by reason of the user not remembering the name of the station of a destination clearly, or the method of voice input not being known), and you may make it switch to a touch input. Of course, after displaying an initial screen format again and calling a user's attention, you may make it switch to a touch input once.

[0054] Now, a user will do the touch input of the initial "**" of the name of the station "Asakusabashi" to wish, if the screen shown in drawing 10 is displayed (step 712). Then, the list for a touch input to which an initial as shows a control section 13 to drawing 11 makes the name of the station of "**" an input item is displayed (step 713). [0055] Then, a user does the touch input of "Asakusabashi" (step 714). The number input area 502 in drawing 5 and the same number-of-sheets input area (it omits by a diagram) are secured in the list screen of drawing 11 , and by this field up, the number of sheets of a ticket and the touch input which is "child (or an adult/child) further are possible, and it has come.

[0056] The touch input result by the user is displayed on the same screen as drawing 9 , and if it is as a result of [as desired] a user, a user will touch a "check" carbon button. The count result of the count of "correction" is cleared by the touch of a this "check" carbon button. Moreover, the count result of the count of this "correction" may be made to be cleared by switch to the above-mentioned touch input.

[0057] In addition, a screen display (it displays at once, and in not going out, a screen is switched by page [degree] assignment and it displays)

of the list of all names of the stations as shown in drawing 12 is carried out, and it may be made to consider instead of switching to a touch input as mentioned above as the help which a user inputs with voice. This method is suitable when the user does not remember the name of the station of a destination clearly.

[0058]

[Effect of the Invention] As explained in full detail above, according to this invention, it is suitable for choosing a desired item out of many items, and inputting, but Although it is not suitable for the selection input of the item out of a voice input means by which the content of an input is hard to be recognized correctly for some users, and many items Since cooperation of both [these] the input means can be taken so that the features of the touch input means which can be inputted simply and certainly may be employed efficiently, without being based on a user in case it chooses from a small number of items, a user's alter operation can carry out easily and certainly, and can improve user-friendliness.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The block diagram showing the outline configuration of the multi-modal equipment concerning the 1st example of this invention.

[Drawing 2] Drawing showing a part of flow chart for explaining an example of the utilization procedure in this example.

[Drawing 3] Drawing showing the remainder of the flow chart for explaining an example of the utilization procedure in this example.

[Drawing 4] Drawing showing the example of a screen display at the time of classifying and displaying many items in this example.

[Drawing 5] Drawing showing the example of the display screen at the time of displaying greatly the item belonging to one classification chosen by the touch input in this example.

[Drawing 6] Drawing showing the example of the display screen at the time of displaying greatly the item belonging to one classification chosen by voice input in this example.

[Drawing 7] The flow chart for explaining an example of the utilization procedure in the 2nd example of this invention.

[Drawing 8] Drawing showing an example of the initial screen format in this example.

[Drawing 9] Drawing showing an example of the recognition result display screen in this example.

[Drawing 10] Drawing showing an example of the screen for a touch input in this example.

[Drawing 11] Drawing showing an example of the screen for a touch input which makes the lower layer of the screen of drawing 10.

[Drawing 12] Drawing showing an example of the screen for voice input in this example.

[Description of Notations]

11 [-- The speech recognition section, 15 / -- The touch input section, 16 / -- The voice output section, 17 / -- Access detection section.] -
- The voice input section, 12 -- A display, 13 -- A control section, 14

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

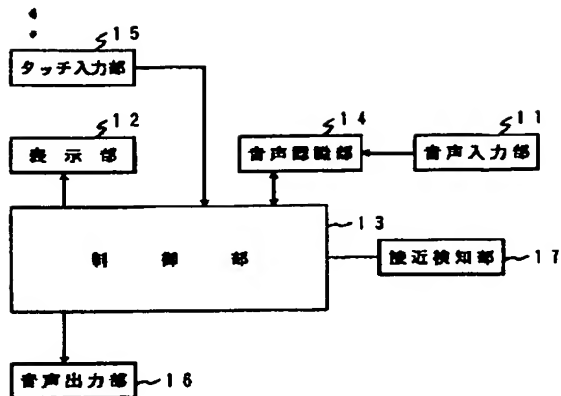
1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DRAWINGS

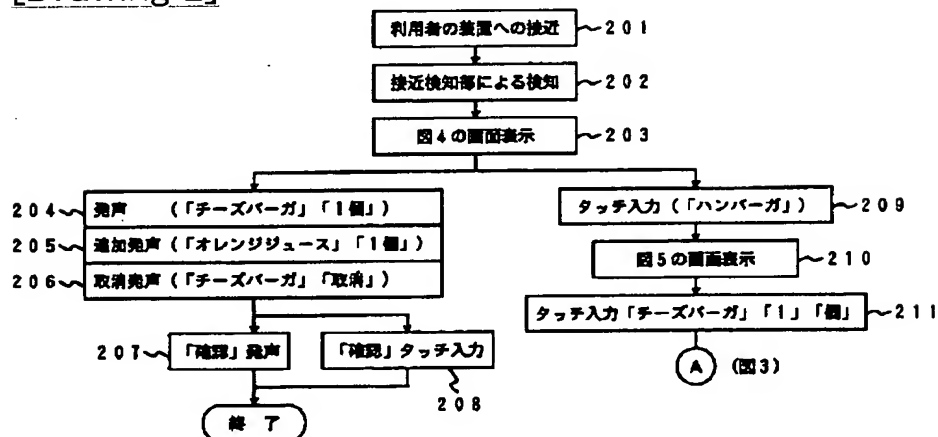
[Drawing 1]



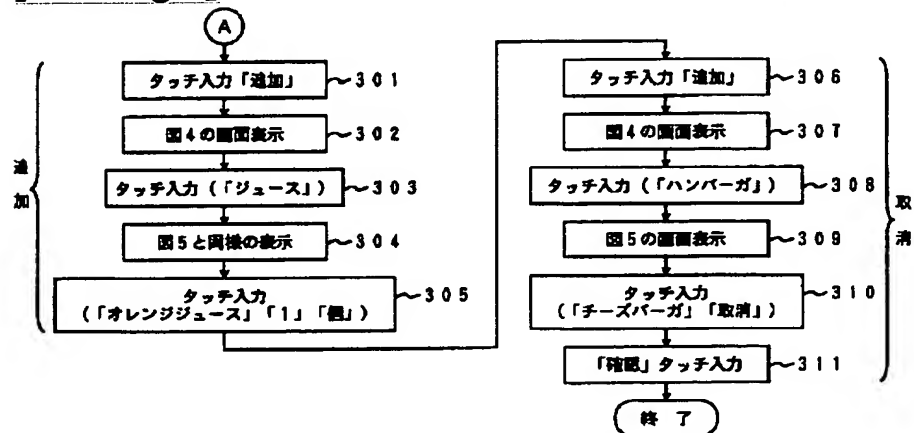
[Drawing 8]



[Drawing 2]



[Drawing 3]



[Drawing 4]

いらっしゃいませ。ご注文をマイクへどうぞ。

ハンバーガー

ハンバーガー	¥390
チーズバーガー	¥390
チリヤキバーガー	¥500
ビッグバーガー	¥600
ヤサイバーガー	¥400
チキンバーガー	¥450

ジュース

オレンジジュース	¥250
グレープジュース	¥250
アップルジュース	¥250
パインジュース	¥250
レモンジュース	¥250

スナック

ホットケーキ	¥300
フライド・ポテトS	¥200
フライド・ポテトL	¥300

クールドリンク

ウーロン茶	¥200
アイスコーヒー	¥220
レモンティー	¥200
ミルクティー	¥200
コーラ	¥200

カツカレー ¥600

ホットドリンク

コーヒー	¥200
レモンティー	¥200
ミルクティー	¥200

＜ご注文＞ ほかにご注文があれば確認をフれて下さい。
 チーズバーガー 1個

確認 407

[Drawing 9]

ご希望の切符は
浅草橋大人1枚
 でよろしいですか。
 よろしければ「確認」にフれて下さい。
 間違っていた場合は「訂正」にフれて下さい。

訂正 901

確認 902

[Drawing 5]

品名をフれたあと、個数をフれて下さい。

ハンバーガー

品名	値段
ハンバーガー	¥390
チーズバーガー	¥390
チリヤキバーガー	¥500
ビッグバーガー	¥600
ヤサイバーガー	¥400
チキンバーガー	¥450

502

1	2	3
4	5	6
7	8	9
0	取消	個

503

＜ご注文＞ ほかにご注文があれば確認をフれて下さい。
 チーズバーガー 1個

追加 確認 505

504

[Drawing 6]

「オレンジジュース」「1個」のようにマイクへどうぞ。

ジュース	
品 名	ね だ ん
オレンジジュース	¥250
グレープジュース	¥250
アップルジュース	¥250
パイナップルジュース	¥250
レモンジュース	¥250

<ご注文>

[Drawing 10]

駅名の漢文字にフレて下さい。

あ	か	さ	た	な	は	ま	や	ら	わ
い	り	
う	る	
え	れ	
お	ろ	

[Drawing 11]

希望の駅名にフレて下さい。

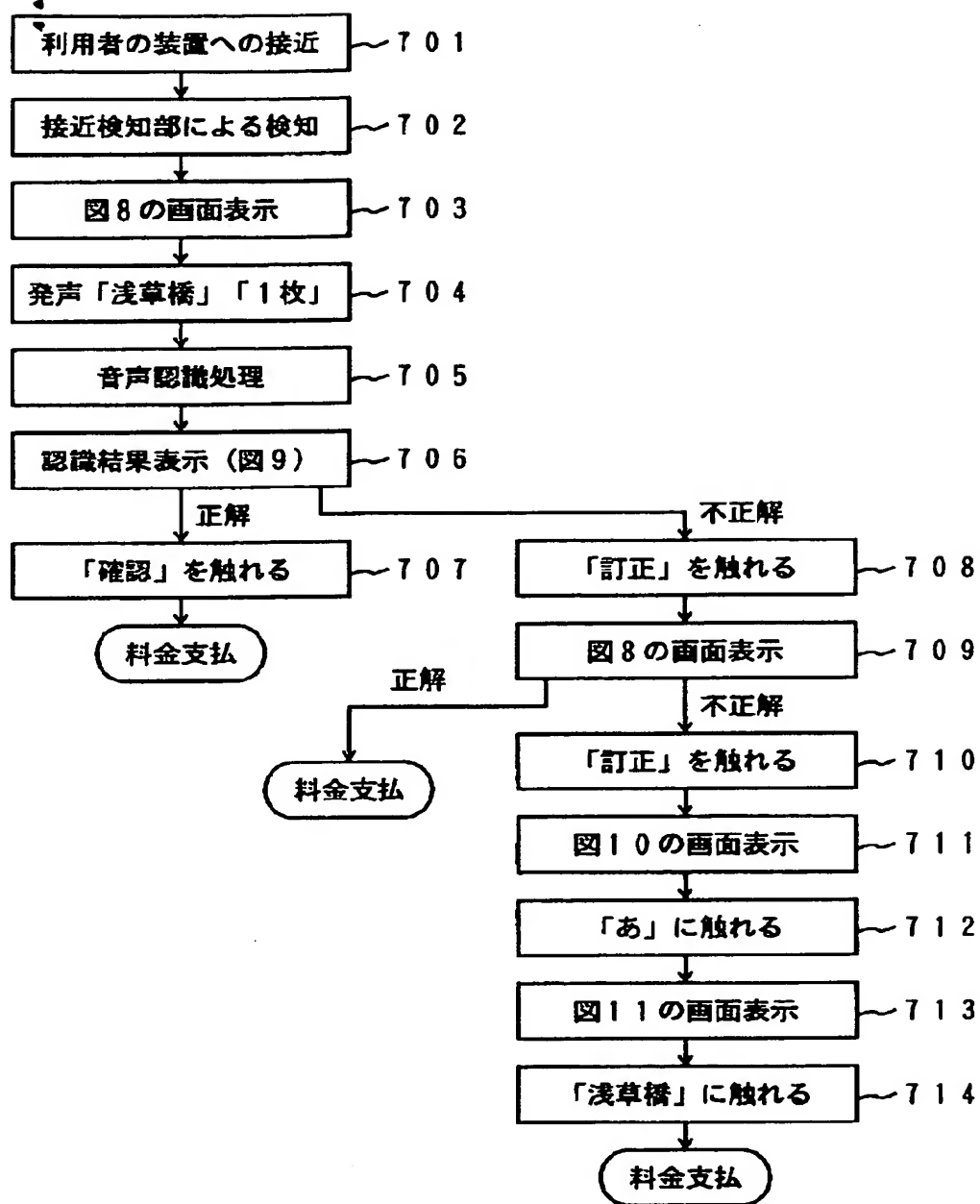
相 原	上 尾
青 場	阿佐ヶ谷
赤 塚	浅草橋
赤 羽	浅 野
秋 川	旭
昭 島	足 尾
秋葉原	次ページ

[Drawing 12]

希望の駅名をマイクへどうぞ。

あ	相 原	530
...
い	原田橋	140
...
...
と	東 京	120

[Drawing 7]



[Translation done.]